Applied Statistical Regression – AS 2014

People:

Lecturer: **Dr. Marcel Dettling** (marcel.dettling@math.ethz.ch)

Coordinators: **Preetam Nandy** (nandy@stat.math.ethz.ch)
Emilija Perkovic (perkovic@stat.math.ethz.ch)

Course Schedule:

All lectures will be held at HG D1.1, on Mondays from 8.15-9.00, resp. 9.15-10.00.

Week	Date	L/E	Topics
01	15.09.2014	---	---
02	22.09.2014	L/L	**Linear Modeling & Smoothing**
03	29.09.2014	E/E	**Introduction to R**
04	06.10.2014	L/L	**Simple Linear Regression**
05	13.10.2014	L/E	**Curvilinear Models, Variable Transformations**
06	20.10.2014	L/L	**Multiple Linear Regression: Fitting and Inference**
07	27.10.2014	L/E	**Extensions: Categorical Variables, Interactions**
08	03.11.2014	L/L	**Model Diagnostics: Residual Plots**
09	10.11.2014	L/E	**Model Choice: Variable Selection**
10	17.11.2014	L/L	**Cross Validation, Modeling Strategies**
11	24.11.2014	L/E	**Logistic and Binomial Regression**
12	01.12.2014	L/L	**Regression for Nominal and Ordinal response**
13	08.12.2014	L/E	**Poisson Regression for Count Data**
14	15.12.2014	L/L	**Advanced Topics**

Exercise Schedule:

The exercises start on September 30, 2013 from 8.15 to 10.00 with an introduction to the statistical software package R. This takes place at the computer labs, the rooms will be communicated by the coordinators via e-mail. Then, the exercise schedule is as follows:

Series	Date	Topic	Hand-In	Discussion
01	29.09.2014	Data Analysis with R	---	29.09.2014
02	29.09.2014	Simple Regression	06.10.2014	13.10.2014
03	13.10.2014	Multiple Regression 1	20.10.2014	27.10.2014
04	27.10.2014	Multiple Regression 2	03.11.2014	10.11.2014
05	10.11.2014	Multiple Regression 3	17.11.2014	24.11.2014
06	24.11.2014	Logistic Regression	01.12.2014	08.12.2014
07	08.12.2014	Count and Ordinal Data	---	08.12.2014

All exercises except the R introduction take place at HG D1.1 (group of Perkovic) and HG D3.2 (group of Nandy). All students whose last name starts with letters A-K visit the group of Perkovic, whereas the ones with letters L-Z visit the Nandy group.

The solved exercises should be handed in at the end of the lecture of the due date or placed in the corresponding tray in HG J68 until 12.00am. Please note that only solutions to the exercises with your most important findings and answers shall be handed in, but no R script files and lengthy compilations of output or figures!

Software:

Theory and exercises will be based on the statistical software R. This is a freely available open source program that works on all platforms and has become worldwide standard for data analysis. It can be downloaded from CRAN (http://cran.r-project.org/). A good primer is the R tutorial that will be discussed in first exercise session of September 29, 2014. Further documentation on the use of R is available on CRAN, and in many textbooks.

Exercises and Exam:

There will be a written exam during the regular session that lasts for 120 minutes. It will be "open book", thus you are allowed to bring any written materials you wish. We also recommend bringing a pocket calculator. However, notebooks/computers are not allowed. In this regard, there will be some discrepancy between the exam and the exercises. The latter are mostly with R and try to best prepare you for applying regression analyses in scientific context. However, the paper and pencil exam will at best address your competency in applied regression and not be based on theory only.

Furthermore, the exercises will predominantly address topics that were discussed in previous lectures. However, because exercises are bi-weekly only, there may be some cases where topics of the subsequent lecture will also be covered. Also, the exercises tend to contain several problems and may seem rather long. This is because they serve as a very important illustration to the theory that is presented in the lectures and thus, they are at the heart of learning how to work with regression analysis in practice. Furthermore, all exercises will be complemented by a detailed and commented sample solution. Because there are no conditions for obtaining the attendance certificate, it is completely up to you how many of the problems you want to solve. However, for becoming proficient in applying statistical regression techniques, thoroughly studying the exercises is key.

Written Material

There is a scriptum for this course. The scriptum, as well as the slides that are presented during the lectures, the exercise sheets, the sample solutions and some instructional datasets are also available for download from the course website at http://stat.ethz.ch/education/semesters/as2014/asr.

Credit Points:

There are no conditions for obtaining the attendance certificate. This also holds for PhD students that are after ETH credit points. However, anybody who is after ECTS credit points needs to attend and pass the exam for getting the credits awarded

Literature:

1) **Linear Models with R**, Julian J. Faraway, Chapman & Hall/CRC (2005). ISBN-10: 1584884258. 229 pages, ca. 70$. There is a freely available version on CRAN, that is almost identical to the book, entitled **Practical Regression and Anova using R**: see http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf. For covering the topics from week 11 on, the second volume of Faraway's regression literature is required:

   Extending the Linear Model with R, Julian J. Faraway, Chapman & Hall/CRC (2006). ISBN-10: 158488424X. 312 pages, ca. 75$. There is no free online version.

2) **Applied Regression Analysis**, N. Draper and H. Smith, Wiley Interscience, 3rd Edition (1998). ISBN-10: 0471170828. 736 pages, ca. 100$.

3) **Introduction to Linear Regression Analysis**, D. Montgomery, E. Peck, G. Vining, Wiley-Interscience, 4th Edition (2006). ISBN-10: 0471754951. 640 pages, ca. 85$.

4) **Applied Regression Analysis and Generalized Linear Models**, J. Fox, Sage Publications, 2nd Edition (2008). ISBN-10: 0761930426. 688 pages, ca. 82$.